CORE Metadata, citation and similar papers at core.ac.uk
Provided by Elsevier - Publisher Connector

Journal of the Egyptian Mathematical Society (2016) **24**, 672–680

Egyptian Mathematical Society
Journal of the Egyptian Mathematical Society
www.etms-eg.org
www.elsevier.com/locate/joems

Original Article

An interior-point penalty active-set trust-region algorithm

Bothina El-Sobky*

Assistant Professor in Department of Mathematics and Computer Science, Faculty of Science, Alexandria University, Egypt

Received 17 December 2015; revised 25 March 2016; accepted 20 April 2016
Available online 24 May 2016

Keywords
Active set;
Penalty method;
Interior-point;
Coleman–Li strategy;
Trust region;
Global convergence

Abstract In this work, an active set strategy is used together with a Coleman–Li strategy and penalty method to transform a general nonlinear programming problem with bound on the variables to unconstrained optimization problem with bound on the variables. A trust-region globalization strategy is used to compute a step. A global convergence theory for the proposed algorithm is presented under credible assumptions.

Prefatory numerical experiment on the algorithm is presented. The rendering of the algorithm is reported on some classical problem.

2010 Mathematics Subject Classification: 90C30; 90C55; 65K05; 49M37,

Copyright 2016, Egyptian Mathematical Society. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

In this paper, we consider the following constrained optimization problem

$$\begin{array}{lllll} \textit{minimize} & f(x) & & & \\ \textit{subject to} & a_i(x) & = & 0 & i \in E, \\ & a_i(x) & \le & 0 & i \in I, \\ & \alpha \le x \le \beta, & & & \end{array} \tag{1.1}$$

where $f: \Re^n \to \Re$, $a_i: \Re^n \to \Re^m$, $E \bigcup I = \{1, ..., m\}$ and $E \bigcap I = \emptyset$, $\alpha \in \{\Re \bigcup \{-\infty\}\}^n$, $\beta \in \{\Re \bigcup \{\infty\}\}^n$, $m < n$, and $\alpha < \beta$. The functions f and a_i, $i = \{1, ..., m\}$ are presumed to be at least twice continuously differentiable. We denote the feasible set $\boldsymbol{F} = \{x : \alpha \le x \le \beta\}$ and the strict interior feasible set $int(\boldsymbol{F}) = \{x : \alpha < x < \beta\}$.

In this paper, we use an active-set strategy in [1] to convert the above problem to an equality constrained optimization problem with bounded variables. The head feature of the suggested active set is that it is identified and updated naturally by the step. See [2–4].

A penalty method is used in this paper to transform the equality constrained optimization problem which was obtained from the above step to unconstrained optimization problem with bound on variables. Some penalty functions have been suggested and many contributions addressing the convergence of these methods have been made, see [5,6].

* Corresponding author. Tel.: +201156646045.
E-mail address: bothinaelsobky@yahoo.com
Peer review under responsibility of Egyptian Mathematical Society.

Production and hosting by Elsevier

S1110-256X(16)30025-6 Copyright 2016, Egyptian Mathematical Society. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
http://dx.doi.org/10.1016/j.joems.2016.04.003

A Coleman–Li strategy in [7] is used to form a sequential quadratic programming subproblem of unconstrained optimization problem. For more details, see [7–9].

In this paper, we use a trust-region strategy to evaluate a step. A trust-region strategy is globalization method which means modifying the local method in such a way that it is ensured to converge even if the starting point is far away from the solution. Most trust-region algorithms for solving a constrained optimization problem try to merge the trust-region idea with the sequential quadratic programming method. See [2–4,10]. Under credible assumptions, a convergence theory for our algorithm is introduced.

The rest of this section introduces some notations that are used throughout the rest of the paper. The paper is arranged as follows. In Section 2, a detailed characterization of the main steps to form a sequential quadratic programming subproblem is introduced. In Section 3, a detailed characterization of an interior-point trust-region algorithm is given. Sections 4–9 are devoted to the global convergence theory of the proposed algorithm under important assumptions. Section 10 contains a Matlab implementation of the interior-point trust-region algorithm and our numerical results. Finally, Section 11 contains concluding remarks.

In this paper, we use the symbol $f_k = f(x_k)$, $\nabla f_k = \nabla f(x_k)$, $\nabla^2 f_k = \nabla^2 f(x_k)$, $A_k = A(x_k)$, $\nabla A_k = \nabla A(x_k)$, $Z_k = Z(x_k)$, $W_k = W(x_k)$ and so on to denote the function value at a particular point. We denote to the Hessian of the objective function f_k or an approximation to it by H_k . Finally, all norms are l_2-norms.

2. A sequential quadratic subproblem

Motivated by the active-set strategy in [1], we define a 0–1 diagonal matrix $W(x) \in \Re^{m\times m}$ whose diagonal entries are

$$w_i(x) = \begin{cases} 1, & \text{if } i \in E, \\ 1, & \text{if } i \in I \text{ and } a_i(x) \geq 0, \\ 0, & \text{if } i \in I \text{ and } a_i(x) < 0. \end{cases} \tag{2.1}$$

Using the above matrix, problem (1.1) is converted to the following

$$\begin{array}{ll} \textit{minimize} & f(x), \\ \textit{subject to} & A(x)^T W(x) A(x) = 0, \\ & \alpha \leq x \leq \beta, \end{array}$$

where $A(x) = (a_1(x), \ldots, a_m(x))^T$ is a continuously differentiable function.

Using a penalty method, the above problem is transformed to the following unconstrained optimization problem with bounds on the variable

$$\begin{array}{ll} \textit{minimize} & f(x) + \dfrac{r}{2}\|W(x)A(x)\|^2, \\ \textit{subject to} & \alpha \leq x \leq \beta, \end{array} \tag{2.2}$$

where $r > 0$ is a penalty parameter. Let

$$\phi(x; r) = f(x) + \frac{r}{2}\|W(x)A(x)\|^2. \tag{2.3}$$

The Lagrangian function associated with bounded problem (2.2) is given by

$$L(x, \lambda, \mu; r) = \phi(x; r) - \lambda^T(x - \alpha) - \mu^T(\beta - x), \tag{2.4}$$

where λ and μ are Lagrange multiplier vectors associated with the inequality constraints $x - \alpha \geq 0$ and $\beta - x \geq 0$ respectively.

The first-order necessary conditions for a point x_* to be a solution of problem (1.1) are the existence of multipliers $\lambda_* \in \Re^n_+$, and $\mu_* \in \Re^n_+$, such that (x_*, λ_*, μ_*) satisfies

$$\nabla\phi(x_*; r_*) - \lambda_* + \mu_* = 0, \tag{2.5}$$

$$\alpha \leq x_* \leq \beta, \tag{2.6}$$

and for all j corresponding to $x^{(j)}$ with finite bound, we have

$$\lambda_*^{(j)}(x_*^{(j)} - \alpha^{(j)}) = 0, \tag{2.7}$$

$$\mu_*^{(j)}(\beta^{(j)} - x_*^{(j)}) = 0, \tag{2.8}$$

where $\nabla\phi(x_*; r_*) = \nabla f(x_*) + r_*\nabla A(x_*)W(x_*)A(x_*)$.

Let $Z(x)$ be the diagonal scaling matrix whose diagonal elements are given by

$$z^{(j)}(x) = \begin{cases} \sqrt{(x^{(j)} - \alpha^{(j)})}, & \text{if } (\nabla\phi(x; r))^{(j)} \geq 0 \text{ and } \alpha^{(j)} > -\infty, \\ \sqrt{(\beta^{(j)} - x^{(j)})}, & \text{if } (\nabla\phi(x; r))^{(j)} < 0 \text{ and } \beta^{(j)} < +\infty, \\ 1, & \text{otherwise.} \end{cases} \tag{2.9}$$

For more details see [7,8].

Using the diagonal scaling matrix $Z(x)$, the first order necessary conditions for the point x_* to solve problem (1.1) are that $x_* \in \boldsymbol{F}$ and solves the following nonlinear system

$$Z^2(x)\nabla\phi(x; r) = 0. \tag{2.10}$$

Any point $x_* \in \boldsymbol{F}$ that satisfies the condition (2.10) is called a Karush–Kuhn–Tucker point or KKT point. For more details see [5].

A system (2.10) is continuous but not differentiable at some point $x \in \boldsymbol{F}$. The non-differentiability happens when $z^{(j)} = 0$ and these points are averted by restricting $x \in int\boldsymbol{F}$. Also the non-differentiability happens when a variable $x^{(j)}$ has a finite lower bound and an infinite upper bound and $(\nabla\phi(x; r))^{(j)} = 0$. But these points are not significant, so we define a vector $\psi(x)$ whose components are $\psi^{(j)}(x) = \frac{\partial((z^{(j)})^2)}{\partial x^{(j)}}$, $j = 1, \ldots, n$ such that $\psi^{(j)}$ to be zero whenever $(\nabla\phi(x; r))^{(j)} = 0$. Hence, we can write

$$\psi^{(j)}(x) = \begin{cases} 1, & \text{if } (\nabla\phi(x; r))^{(j)} \geq 0 \text{ and } \alpha^{(j)} > -\infty, \\ -1, & \text{if } (\nabla\phi(x; r))^{(j)} < 0 \text{ and } \beta^{(j)} < +\infty, \\ 0, & \text{otherwise.} \end{cases} \tag{2.11}$$

Assuming $x \in int(\boldsymbol{F})$ and applied Newton's method on the system (2.10), then we have

$$\begin{aligned} &[Z^2(x)\nabla^2\phi(x; r) + diag(\nabla\phi(x; r))diag(\psi(x))]\Delta x \\ &\quad = -Z^2(x)\nabla\phi(x; r), \end{aligned} \tag{2.12}$$

where

$$\nabla^2\phi(x; r) = H + r\nabla A(x)W(x)\nabla A(x)^T, \tag{2.13}$$

and H is the Hessian of the objective function $f(x)$ or an approximation to it. Multiplying both sides of Eq. (2.12) by $Z^{-1}(x)$ and scale the step using $\Delta x = Z(x)s$, then we have

$$\begin{aligned} &[Z(x)\nabla^2\phi(x; r)Z(x) + diag(\nabla\phi(x; r))diag(\psi(x))]s \\ &\quad = -Z(x)\nabla\phi(x; r), \end{aligned} \tag{2.14}$$

Notice that (2.14) represents the first order necessary condition of the following sequential quadratic programming problem

$$minimize \quad \phi(x;r) + (Z\nabla\phi(x;r))^T s + \frac{1}{2}s^T Bs, \tag{2.15}$$

where

$$B = Z(x)\nabla^2\phi(x;r)Z(x) + diag(\nabla\phi(x;r))diag(\psi(x)). \tag{2.16}$$

That is, the point x_* that satisfies the first order necessary condition of problem (2.15) will satisfy the first order necessary condition of problem (1.1).

In the following section, we present main steps of our interior-point trust-region algorithm for solving problem (1.1).

3. An interior-point trust-region algorithm

This section is devoted to the description of a new interior-point method.

3.1. Evaluating a step s_k

In this section, a step s_k is computed by solving the following trust-region subproblem

$$\begin{array}{ll} minimize & q_k(Z_k s) = \phi(x_k;r_k) + (Z_k\nabla\phi(x_k;r_k))^T s + \frac{1}{2}s^T B_k s \\ subject\ to & \| s \| \leq \delta_k, \end{array} \tag{3.1}$$

where $\delta_k > 0$ is the radius of the trust region.

It is not necessary to obtain a very precise approximation to the solution of the subproblem (3.1). Instead any approximation to the solution of the subproblem (3.1) can be used as long as the predicted decrease obtained by s_k is greater than or equal to a fraction of the predicted decrease obtained by the Cauchy step s_k^{cp}. That is

$$q_k(0) - q_k(Z_k s_k) \geq \varphi[q_k(0) - q_k(Z_k s_k^{cp})], \tag{3.2}$$

for some $\varphi \in (0,1]$, where s_k^{cp} is given by

$$s_k^{cp} = -t_k^{cp}(Z_k\nabla\phi(x_k;r_k)),$$

and the parameter t_k^{cp} is defined by

$$t_k^{cp} = \begin{cases} \frac{\|(Z_k\nabla\phi(x_k;r_k))\|^2}{(Z_k\nabla\phi(x_k;r_k))^T B_k (Z_k\nabla\phi(x_k;r_k))} & \text{if } \frac{\|Z_k\nabla\phi(x_k;r_k)\|^3}{(Z_k\nabla\phi(x_k;r_k))^T B_k (Z_k\nabla\phi(x_k;r_k))} \\ & \leq \delta_k \text{ and } (Z_k\nabla\phi(x_k;r_k))^T \\ & B_k(Z_k\nabla\phi(x_k;r_k)) > 0, \\ \frac{\delta_k}{\|Z_k\nabla\phi(x_k;r_k)\|} & \text{otherwise.} \end{cases} \tag{3.3}$$

Therefore, a generalized dogleg algorithm introduced by Steihaug [11] can be used to evaluate the step. Once the step s_k is evaluated, the damping parameter τ_k is computed to ensure that $x_{k+1} \in int(\boldsymbol{F})$. Our way of evaluating the damping parameter τ_k is presented in Step 4 of algorithm (3.3) below.

Another damping parameter θ_k may be needed to satisfy $x_k \in int(\boldsymbol{F})$, where θ_k is defined as follows. If $(x_k + \tau_k Z_k s_k) \in int(\boldsymbol{F})$, we set $\theta_k = 1$. Otherwise, we set $x_{k+1} = x_k + \theta_k\tau_k Z_k s_k$, such that $\theta_k \in [1 - \sigma\|Z_k s_k\|, 1]$ and $\sigma > 0$ is a pre-specified fixed constant. It is easy to see that $1 - \theta_k = O(\|Z_k s_k\|)$.

3.2. Accepting s_k and updating δ_k

After s_k is obtained, the penalty function $\phi(x_k;r_k)$ is used as a merit function to test if the step s_k is accepted or not. This is done by comparing $Pred_k$ against $Ared_k$. The actual reduction $Ared_k$ is defined as

$$Ared_k = \phi(x_k;r_k) - \phi(x_k + Z_k\tilde{\tau}_k s_k;r_k)$$

where $\tilde{\tau}_k = \theta_k\tau_k$. $Ared_k$ can be written as

$$\begin{aligned} Ared_k &= f(x_k) - f(x_k + Z_k\tilde{\tau}_k s_k) \\ &\quad + \frac{r_k}{2}[\| W_k A_k \|^2 - \| W_{k+1}A_{k+1} \|^2]. \end{aligned} \tag{3.4}$$

The predicted reduction $Pred_k$ is defined to be

$$\begin{aligned} Pred_k &= -(Z_k\nabla f_k)^T\tilde{\tau}_k s_k - \frac{1}{2}\tilde{\tau}_k^2 s_k^T G_k s_k \\ &\quad + \frac{r_k}{2}[\| W_k A_k \|^2 - \| W_k(A_k + (Z_k\nabla A_k)^T\tilde{\tau}_k s_k) \|^2], \end{aligned} \tag{3.5}$$

where

$$G_k = Z_k H_k Z_k + diag(\nabla\phi(x_k;r_k))diag(\eta_k).$$

Our way of testing s_k and updating the trust-region radius δ_k is presented in the following algorithm.

Algorithm 3.1 (Test s_k and update the trust-region radius algorithm). Choose $0 < \eta_1 < \eta_2 \leq 1$, $\delta_{max} > \delta_{min}$, and $0 < \alpha_1 < 1 < \alpha_2$.

While $\frac{Ared_k}{Pred_k} < \eta_1$, or $Pred_k \leq 0$.
Set $\delta_k = \alpha_1 \| s_k \|$.
Evaluate a new s_k.

If $\eta_1 \leq \frac{Ared_k}{Pred_k} < \eta_2$, then set $x_{k+1} = x_k + \tilde{\tau}_k Z_k s_k$.
$\delta_{k+1} = \max(\delta_k, \delta_{min})$.
End if.
If $\frac{Ared_k}{Pred_k} \geq \eta_2$, then set $x_{k+1} = x_k + \tilde{\tau}_k Z_k s_k$.
$\delta_{k+1} = \min\{\delta_{max}, \max\{\delta_{min}, \alpha_2\delta_k\}\}$.
End if.

To update the penalty parameter r_k, we use a scheme suggested by Yuan [12]. Our way of updating r_k is presented in the following algorithm.

Algorithm 3.2 (Update r_k algorithm).

Set $r_0 = 1$. Compute $Pred_k$ given by Eq. (3.5).
If

$$Pred_k \geq \| Z_k\nabla A_k W_k A_k \| \min\{\| Z_k\nabla A_k W_k A_k \|, \delta_k\}. \tag{3.6}$$

Set $r_{k+1} = r_k$.
Else, set $r_{k+1} = 2r_k$.

End if

Either $\| Z_k \nabla f_k \| + \| Z_k \nabla A_k W_k A_k \| \leq \epsilon_1$ or $\|s_k\| \leq \epsilon_2$ for some $\epsilon_1 > 0$ and $\epsilon_2 > 0$, the algorithm is stopped

3.3. The master algorithm

Master steps of our algorithm is presented in the following algorithm.

Algorithm 3.3 (The Main Algorithm)**.** Step 0.

Given $x_0 \in int(\boldsymbol{F})$. Compute matrices W_0, Z_0 and ψ_0. Set $r_0 = 1$.
Choose $\epsilon_1, \epsilon_2, \alpha_1, \alpha_2, \eta_1, \eta_2$ and σ such that $\epsilon_1 > 0, \epsilon_2 > 0$, $0 < \alpha_1 < 1 < \alpha_2$, $0 < \eta_1 < \eta_2 \leq 1$, and $\sigma > 0$. Choose δ_{min}, δ_{max}, and δ_0 such that $\delta_{min} \leq \delta_0 \leq \delta_{max}$.
Set $k = 0$.

Step 1. If $\| Z_k \nabla f_k \| + \| Z_k \nabla A_k W_k A_k \| \leq \epsilon_1$, then stop.
Step 3. A step s_k is evaluated by solving the subproblem (3.1).
If $\|s_k\| \leq \epsilon_2$, stop, end.
Step 4.

(a) Compute

$$u_k^{(i)} = \begin{cases} \frac{\alpha^{(i)} - x_k^{(i)}}{Z_k^{(i)} s_k^{(i)}}, & \text{if } \alpha^{(i)} > -\infty \text{ and } Z_k^{(i)} s_k^{(i)} < 0 \\ 1, & \text{otherwise,} \end{cases}$$

(b) Compute

$$v_k^{(i)} = \begin{cases} \frac{\beta^{(i)} - x_k^{(i)}}{Z_k^{(i)} s_k^{(i)}}, & \text{if } \beta^{(i)} < \infty \text{ and } Z_k^{(i)} s_k^{(i)} > 0 \\ 1, & \text{otherwise.} \end{cases}$$

(c) Compute

$$\tau_k = \min\{1, \min_i \{u_k^{(i)}, v_k^{(i)}\}\}. \tag{3.7}$$

(d) Set $x_{k+1} = x_k + \tau_k Z_k s_k$.
If $x_{k+1} \in int(\boldsymbol{F})$, then go to step 5.
Else, set $x_{k+1} = x_k + \theta_k \tau_k Z_k s_k$, end.

Step 5. Compute W_{k+1} given by (2.1).
Step 6. Test the step and update the trust-region radius using algorithm (3.1).
Step 7. Update the penalty parameter r_k using algorithm (3.2).
Step 8. Compute Z_{k+1} given by (2.9) and ψ_{k+1} given by (2.11).
Step 9. Set $k = k + 1$ and go to Step 1.

In Sections 5–9, we present our global convergence theory to the set of the first order points under some assumptions which are stated in the following section.

4. General assumptions

Let $\{x_k\}$ be the sequence of points generated by the algorithm (3.3) and let Ω be a convex subset of $\Re^n$ that contains all iterates $x_k \in int(\boldsymbol{F})$ and $x_k + \tilde{\tau}_k Z_k s_k \in int(\boldsymbol{F})$, for all trial steps s_k.

On the set Ω, we state the following general assumptions under which our global convergence theory is proved.

A general assumptions:

[GA_1] The functions f and a_i, $i = \{1, ..., m\}$ are presumed to be at least twice continuously differentiable $\forall x \in \Omega$.
[GA_2] All of $f(x)$, $\nabla f(x)$, $\nabla^2 f(x)$, $a_i(x)$, $\nabla a_i(x)$, $i = \{1, ..., m\}$ are uniformly bounded in Ω.
[GA_3] The sequence of Hessian matrices $\{H_k\}$ is bounded.

In the above general assumptions, we do not presume $\nabla a_i(x)$, $i = \{1, ..., m\}$ has inverse for all $x \in \Omega$. So, we may have other kinds of stationary points. They are presented in the following section.

5. Stationary points

In this section, we define four kinds of stationary points, a Fritz John point, an infeasible Fritz John point, an infeasible Mayer–Bliss point, and a KKT point.

Definition 5.1 (Fritz John point)**.** A point $x_* \in \Omega$ is called a Fritz John point, if there exist $\gamma_* \in \Re$ and a Lagrange multiplier vector $\nu_* \in \Re^m$ not all zero such that:

$$\gamma_* Z(x_*) \nabla f(x_*) + Z(x_*) \nabla A(x_*) \nu_* = 0, \tag{5.1}$$

$$W(x_*) A(x_*) = 0, \tag{5.2}$$

$$\gamma_*, \quad (\nu_*)_i \geq 0, \quad i = 1, 2, ..., m, \tag{5.3}$$

The above conditions are called Fritz John conditions. For more details see [13].

If $\gamma_* \neq 0$, then Eqs. (5.1)–(5.3) correspond with KKT conditions (2.10) and the point $(x_*, 1, \frac{\nu_*}{\gamma_*})$ is called KKT point.

Definition 5.2 (Infeasible Fritz John point)**.** A point $x_* \in \Omega$ is called an infeasible Fritz John point, if there exist $\gamma_* \in \Re$ and a Lagrange multiplier vector $\nu_* \in \Re^m$ not all zero such that:

$$\gamma_* Z(x_*) \nabla f(x_*) + Z(x_*) \nabla A(x_*) \nu_* = 0, \tag{5.4}$$

$$Z(x_*) \nabla A(x_*) W(x_*) A(x_*) = 0 \;\; \textit{but} \;\; \|W(x_*) A(x_*)\| > 0, \tag{5.5}$$

$$\gamma_*, \quad (\nu_*)_i \geq 0, \quad i = 1, 2, ..., m. \tag{5.6}$$

The above conditions are called infeasible Fritz John conditions. For more details see [13].

If $\gamma_* \neq 0$ then Eqs. (5.4)–(5.6) are called an infeasible KKT conditions and the point $(x_*, 1, \frac{\nu_*}{\gamma_*})$ is called an infeasible KKT point.

Definition 5.3 (Infeasible Mayer–Bliss point)**.** A point $x_* \in \Omega$ is called an infeasible Mayer–Bliss point if

$$Z(x_*) \nabla A(x_*) W(x_*) A(x_*) = 0,$$
$$\|W(x_*) A(x_*)\| > 0.$$

The above conditions are called infeasible Mayer–Bliss conditions. For more details see [14].

The conditions stated in Definitions (5.1)–(5.3) are called stationary conditions of problem (1.1) and the point that satisfies any of these stationary conditions is called a stationary point.

The following three lemmas provide conditions equivalent to the conditions given in Definitions (5.1)–(5.3)

Lemma 5.1. *Suppose GA_1–GA_3. A subsequence $\{x_{k_i}\}$ of the iteration sequence asymptotically satisfies infeasible Mayer–Bliss conditions if it satisfies:*

(1) $\lim_{k_i\to\infty} \|W_{k_i}A_{k_i}\| > 0$.

(2) $\lim_{k_i\to\infty} \{\min_{s\in\Re^{n-m}} \{\|W_{k_i}(A_{k_i} + (Z_k\nabla A_{k_i})^T \tilde{\tau}_{k_i} s)\|^2\}\} = \lim_{k_i\to\infty} \|W_{k_i}A_{k_i}\|^2$.

Proof. Let the subsequence $\{k_i\}$ be renamed to $\{k\}$ to simplify the notations avoiding double indices. The minimizer $\hat{s}_k$ of $\min_s \|W_k(A_k + (Z_k\nabla A_k)^T\tilde{\tau}_k s)\|^2$ satisfies

$$\tilde{\tau}_k^2 Z_k\nabla A_k W_k \nabla A_k^T Z_k \hat{s}_k + \tilde{\tau}_k Z_k \nabla A_k W_k A_k = 0, \tag{5.7}$$

From condition 2, we have

$$\lim_{k\to\infty} \{\tilde{\tau}_k^2 \hat{s}_k^T Z_k\nabla A_k W_k \nabla A_k^T Z_k \hat{s}_k + 2\tilde{\tau}_k \hat{s}_k^T Z_k\nabla A_k W_k A_k\} = 0. \tag{5.8}$$

We consider two cases:

(i) If $\lim_{k\to\infty} \hat{s}_k = 0$, then from (5.7) we have

$$\lim_{k\to\infty} \{\tilde{\tau}_k Z_k \nabla A_k W_k A_k\} = 0. \tag{5.9}$$

(ii) If $\lim_{k\to\infty} \hat{s}_k \neq 0$, then by multiplying Eq. (5.7) from the left by $2\hat{s}_k^T$ and subtract it from (5.8), we have

$$\lim_{k\to\infty} \{\tilde{\tau}_k^2 \hat{s}_k^T Z_k\nabla A_k W_k \nabla A_k^T Z_k \hat{s}_k\} = 0.$$

But this implies that $\lim_{k\to\infty} \{\tilde{\tau}_k Z_k\nabla A_k W_k A_k\} = 0$. Hence in either case, we have

$$\lim_{k\to\infty} \{Z_k\nabla A_k W_k A_k\} = 0,$$

where $\lim_{k\to\infty} \tilde{\tau}_k = 1$. Thus conditions of Definition (5.3) hold in the limit. □

Lemma 5.2. *Suppose GA_1-GA_3. A subsequence $\{x_{k_i}\}$ of the iteration sequence asymptotically satisfies the infeasible Fritz John conditions if it satisfies:*

(1) $\lim_{k_i\to\infty} \|W_{k_i}A_{k_i}\| > 0$.

(2) $\lim_{k_i\to\infty} Z_{k_i}\nabla A_{k_i} W_{k_i} A_{k_i} = 0$.

Proof. Let the subsequence $\{k_i\}$ be renamed to $\{k\}$ to simplify the notations avoiding double indices. From condition 2, we can write

$$\lim_{k\to\infty} Z_k\nabla A_k W_k A_k = 0.$$

Take $(\nu_k)_i = (W_kA_k)_i$, $i = 1, ..., m$. Since $\lim_{k\to\infty}\|W_kA_k\| > 0$, then $\lim_{k\to\infty}(\nu_k)_i \geq 0$, for $i = 1, ..., m$ and $\lim_{k\to\infty}(\nu_k)_i > 0$, for some i. Therefore $\lim_{k\to\infty} Z_k\nabla A_k \nu_k = 0$. Thus in the limit with $\gamma_* = 0$, the conditions of Definition (5.2) hold . □

If $\lim_{k_i\to\infty} \nabla f_{k_i} + Z_{k_i}\nabla A_{k_i}\nu_{k_i} = 0$, then the infeasible (KKT) conditions are satisfied in the limit. Otherwise, the infeasible Fritz John conditions are satisfied.

Lemma 5.3. *Suppose GA_1-GA_3. A subsequence $\{x_{k_i}\}$ of the iteration sequence asymptotically satisfies the Fritz John's conditions if it satisfies:*

(1) For all k_i, $\|W_{k_i}A_{k_i}\| > 0$ and $\lim_{k_i\to\infty} W_{k_i}A_{k_i} = 0$.

(2) $\lim_{k_i\to\infty} \left\{\min_{s\in\Re^{n-m}} \left\{\frac{\|W_{k_i}(A_{k_i} + (Z_k\nabla A_{k_i})^T\tilde{\tau}_{k_i}s)\|^2}{\|W_{k_i}A_{k_i}\|^2}\right\}\right\} = 1.$

Proof. Let the subsequence $\{k_i\}$ be renamed to $\{k\}$ to simplify the notations avoiding double indices. The limit in Condition 2 is equivalent to

$$\lim_{k\to\infty} \left\{\min_{d\in\Re^n} \{\|U_k + W_k\nabla A_k^T Z_k\tilde{\tau}_k d\|^2\}\right\} = 1, \tag{5.10}$$

where U_k is a unit vector in the direction of W_kA_k, $d = \frac{s}{\|W_kA_k\|}$.

Let $\bar{d}_k$ be a minimizer to the following problem

$$\min_{d\in\Re^n} \{\|U_k + W_k(Z_k\nabla A_k)^T\tilde{\tau}_k d\|^2\}, \tag{5.11}$$

then, from the optimality conditions we have

$$(Z_k\nabla A_k)W_k(Z_k\nabla A_k)^T\tilde{\tau}_k^2\bar{d}_k + (Z_k\nabla A_k)W_kU_k\tilde{\tau}_k = 0, \tag{5.12}$$

We consider two cases:

(i) If $\lim_{k\to\infty} \bar{d}_k = 0$ in the above equation, then we have, $\lim_{k\to\infty}(Z_k\nabla A_k)W_kU_k\tilde{\tau}_k = 0$.

(ii) If $\lim_{k\to\infty} \bar{d}_k \neq 0$, then from (5.10) and the fact that $\bar{d}_k$ is a solution to the minimization problem in (5.11), we have

$$\lim_{k\to\infty} \{\bar{d}_k^{\,T}(Z_k\nabla A_k)W_k(Z_k\nabla A_k)^T\tilde{\tau}_k^2\bar{d}_k + 2U_k^T W_k(Z_k\nabla A_k)^T\tilde{\tau}_k\bar{d}_k\} = 0.$$

Multiplying (5.12) from the left by $2\bar{d}_k^{\,T}$ and subtract it from the above limit, we have

$$\lim_{k\to\infty} \tilde{\tau}_k^2\bar{d}_k(Z_k\nabla A_k)W_k(Z_k\nabla A_k)^T\bar{d}_k = 0.$$

This implies $\lim_{k\to\infty}\{Z_k\nabla A_k W_k U_k\tilde{\tau}_k\} = 0$. Hence in both cases, we have

$$\lim_{k\to\infty} \{Z_k\nabla A_k W_k U_k\} = 0, \tag{5.13}$$

where $\lim_{k\to\infty}\tilde{\tau}_k = 1$. The rest of the proof follows using arguments similar to those in the above lemma. □

From lemma (5.3) we can see that, for any subsequence of the iteration sequence that satisfies Fritz John's conditions, the corresponding subsequence of smallest singular values of $W_k\nabla A_k^T$ is not bounded away from zero. This means that asymptotically the gradients of the active constraints are linear dependent. In the following section, we introduce some lemmas which is needed in the proof of our main results.

6. Important lemmas

The following lemma shows that, at any iteration k, the predicted reduction $Pred_k$ is at least equal to the decrease in the quadratic model of the penalty function obtained by the Cauchy step.

Lemma 6.1. *Suppose GA_1–GA_3. Then for all $k > \bar{k}$, there exists a positive constant K_1 independent of the iterates such that,*

$$Pred_k \geq K_1\tilde{\tau}_k \parallel Z_k\nabla\phi(x_k;r_k)\parallel \min\left\{\delta_k, \frac{\parallel Z_k\nabla\phi(x_k;r_k)\parallel}{\parallel B_k\parallel}\right\}. \tag{6.1}$$

Proof. Since the trial step s_k satisfies the fraction-of-Cauchy decrease condition (3.2). Then we consider two cases:

(i) If $s_k^{cp} = -\frac{\delta_k}{\|Z_k \nabla\phi(x_k;r_k)\|}(Z_k\nabla\phi(x_k;r_k))$ and $\|Z_k\nabla\phi(x_k;r_k)\|^3 \geq \delta_k[(Z_k\nabla\phi(x_k;r_k))^T B_k(Z_k\nabla\phi(x_k;r_k))]$, then

$$\begin{aligned} q_k(0) - q_k(Z_k s_k^{cp}) &= -(Z_k\nabla\phi(x_k;r_k))^T s_k^{cp} - \frac{1}{2} s_k^{cp^T} B_k s_k^{cp} \\ &= \frac{\delta_k}{\| Z_k\nabla\phi(x_k;r_k) \|} \| Z_k\nabla\phi(x_k;r_k) \|^2 \\ &- \frac{1}{2}\frac{\delta_k^2}{\| Z_k\nabla\phi(x_k;r_k) \|^2}((Z_k\nabla\phi(x_k;r_k))^T B_k(Z_k\nabla\phi(x_k;r_k))) \\ &\geq \frac{1}{2}\delta_k \| Z_k\nabla\phi(x_k;r_k) \| . \end{aligned} \tag{6.2}$$

(ii) If $s_k^{cp} = -\frac{\|Z_k\nabla\phi(x_k;r_k)\|^2}{(Z_k\nabla\phi(x_k;r_k))^T B_k(Z_k\nabla\phi(x_k;r_k))}(Z_k\nabla\phi(x_k;r_k))$, and $\|Z_k\nabla\phi(x_k;r_k)\|^3 \leq \delta_k((Z_k\nabla\phi(x_k;r_k))^T B_k(Z_k\nabla\phi(x_k;r_k)))$, then

$$\begin{aligned} q_k(0) - q_k(Z_k s_k^{cp}) &= -(Z_k\nabla\phi(x_k;r_k))^T s_k^{cp} - \frac{1}{2} s_k^{cp^T} B_k s_k^{cp} \\ &= \frac{1}{2}\frac{\| Z_k\nabla\phi(x_k;r_k) \|^4}{(Z_k\nabla\phi(x_k;r_k))^T B_k(Z_k\nabla\phi(x_k;r_k))} \\ &\geq \frac{\| Z_k\nabla\phi(x_k;r_k) \|^2}{2 \| B_k \|}. \end{aligned} \tag{6.3}$$

From inequalities (3.2), (6.2), and (6.3) we have,

$$q_k(0) - q_k(Z_k s_k) \geq K_1 \| Z_k\nabla\phi(x_k;r_k) \| \min \times \left\{ \delta_k, \frac{\| Z_k\nabla\phi(x_k;r_k) \|}{\| B_k \|} \right\}.$$

From the above inequality and the following fact

$$q_k(0) - q_k(Z_k\tilde{\tau}_k s_k) \geq \tilde{\tau}_k[q_k(0) - q_k(Z_k s_k)]$$

where $0 \leq \tilde{\tau}_k \leq 1$, then we have

$$q_k(0) - q_k(Z_k\tilde{\tau}_k s_k) \geq K_1\tilde{\tau}_k \| Z_k\nabla\phi(x_k;r_k) \| \min \times \left\{ \delta_k, \frac{\| Z_k\nabla\phi(x_k;r_k) \|}{\| B_k \|} \right\}.$$

But the predicted reduction which given by (3.5) can be written as follows

$$Pred_k = q(0) - q(Z_k\tilde{\tau}_k s_k).$$

Hence

$$Pred_k \geq K_1\tilde{\tau}_k \| Z_k\nabla\phi(x_k;r_k) \| \min \left\{ \delta_k, \frac{\| Z_k\nabla\phi(x_k;r_k) \|}{\| B_k \|} \right\}.$$

□

Lemma 6.2. *Suppose GA_1 and GA_3. Then $W(x)A(x)$ is Lipschitz continuous in Ω.*

Proof. The proof is similar to the proof of lemma 4.1 of [1]. □

From the above lemma, we conclude that $A(x)^T W(x)A(x)$ is differentiable and $\nabla A(x)W(x)A(x)$ is Lipschitz continuous in Ω.

Lemma 6.3. *At any iteration k, let $D(x_k) \in \Re^{m\times m}$ be a diagonal matrix whose diagonal entries are*

$$(d_k)_i = \begin{cases} 1 & \text{if } (A_k)_i < 0 \text{ and } (A_{k+1})_i \geq 0, \\ -1 & \text{if } (A_k)_i \geq 0 \text{ and } (A_{k+1})_i < 0, \\ 0 & \text{otherwise}, \end{cases} \tag{6.4}$$

where $i = 1, 2, ..., m$. Then

$$W_{k+1} = W_k + D_k. \tag{6.5}$$

Proof. The proof is similar to the proof of lemma 6.2 of [3]. □

Lemma 6.4. *Suppose GA_1–GA_3. At any iteration k, there exists a positive constant K_1 independent of k, such that*

$$\|D_k A_k\| \leq K_2\|s_k\|, \tag{6.6}$$

where $D_k \in \Re^{m\times m}$ is the diagonal matrix whose diagonal entries are defined in (6.4).

Proof. The proof is similar to the proof of lemma 6.3 of [3]. □

Lemma 6.5. *Suppose GA_1–GA_3, then there exists a constant $K_3 > 0$ that does not depend on k, such that*

$$| Ared_k - Pred_k | \leq K_3\tilde{\tau}_k r_k\|s_k\|^2. \tag{6.7}$$

Proof. From Eqs. (3.4) and (6.5) we have

$$\begin{aligned} Ared_k &= f(x_k) - f(x_k + Z_k\tilde{\tau}_k s_k) + \frac{r_k}{2}[A_k^T W_k A_k \\ &- A(x_k + Z_k\tilde{\tau}_k s_k)^T(W_k + D_k)A(x_k + Z_k\tilde{\tau}_k s_k)]. \end{aligned}$$

From the above equation, Eq. (3.5), and using Cauchy–Schwarz inequality, we have

$$\begin{aligned} |Ared_k - Pred_k| \leq \frac{\tilde{\tau}_k^2}{2} &| s_k^T Z_k(\nabla^2 f(x_k) \\ &- \nabla^2 f(x_k + \xi_1 Z_k\tilde{\tau}_k s_k))Z_k s_k | \\ &+ \frac{\tilde{\tau}_k^2}{2} | s_k^T diag(\nabla\phi(x_k;r_k))diag(\psi)s_k | \\ &+ r_k\tilde{\tau}_k | Z_k(\nabla A_k - \nabla A(x_k + \xi_2 Z_k\tilde{\tau}_k s_k))W_k A_k s_k | \\ &+ \frac{r_k\tilde{\tau}_k^2}{2} | s_k^T Z_k[\nabla A_k W_k \nabla A_k^T \\ &- \nabla A(x_k + \xi_2 s_k)W_k\nabla A(x_k + \xi_2 Z_k\tilde{\tau}_k s_k)^T]Z_k s_k | \\ &+ \frac{r_k\tilde{\tau}_k^2}{2}\|D_k A_k\|^2 + r_k\tilde{\tau}_k | Z_k\nabla A(x_k + \xi_2 Z_k\tilde{\tau}_k s_k)D_k A_k s_k | \\ &+ \frac{r_k\tilde{\tau}_k^2}{2} | s_k^T Z_k[\nabla A(x_k + \xi_2 s_k)D_k\nabla A(x_k + \xi_2 Z_k\tilde{\tau}_k s_k)^T]Z_k s_k | \end{aligned}$$

for some ξ_1 and $\xi_2 \in (0, 1)$. From lemma (6.4) and general assumptions, the proof is completed. □

7. Convergence when r_k goes to infinity

In this section, the convergence of the sequence of iteration is studied when the parameter r_k goes to infinity. From algorithm (3.2), we observe that the sequence $\{r_k\}$ goes to infinity only when there exists an infinite subsequence of indices $\{k_i\}$ indexing iterates of acceptable steps that satisfy, for all $k \in \{k_i\}$

$$Pred_k < \|Z_k\nabla A_k W_k A_k\| \min\{\|Z_k\nabla A_k W_k A_k\|, \delta_k\}. \tag{7.1}$$

The following lemma studies the case when $\lim sup_{k\to\infty}\| W_k A_k\| > 0$.

Lemma 7.1. *Suppose GA_1-GA_3. If $r_k \to \infty$, as $k \to \infty$ and there exists a subsequence $\{k_j\}$ of indices indexing iterates that satisfy $\|W_k A_k\| \geq \epsilon_1 > 0$ for all $k \in \{k_j\}$, then a subsequence of the* *sequence of iteration indexed $\{k_j\}$ satisfies the infeasible Mayer–Bliss conditions in the limit.*

Proof. Let the subsequence $\{k_i\}$ be renamed to $\{k\}$ to simplify the notations avoiding double indices. By using a contradiction we prove this lemma. So we presume that there exists no subsequence of the iteration sequence that satisfies the infeasible Mayer–Bliss conditions in the limit. Using Lemma (5.1), we have for all k, $|\ \|W_k A_k\|^2 - \|W_k(A_k + (Z_k\nabla A_k)^T \tilde{\tau}_k s_k)\|^2\ | \geq \varepsilon_1$ for some $\varepsilon_1 > 0$ and from Definition (5.3), we have, $\|Z_k \nabla A_k W_k A_k\| \geq \varepsilon_2$ for some $\varepsilon_2 > 0$.

Since $r_k \to \infty$, then there exists an infinite number of acceptable iterates at which inequality (7.1) holds. We consider two cases:

(i) If $\|W_k A_k\|^2 - \|W_k(A_k + (Z_k\nabla A_k)^T \tilde{\tau}_k s_k)\|^2 \geq \varepsilon_1$, we have

$$r_k\{\|W_k A_k\|^2 - \|W_k(A_k + (Z_k\nabla A_k)^T \tilde{\tau}_k s_k)\|^2\} \geq r_k \varepsilon_1 \to \infty. \tag{7.2}$$

Since $\tilde{\tau}_k \to 1$ as $k \to \infty$, then under assumptions GA_2-GA_3 and using (3.5) and (7.2), we have $Pred_k \to \infty$. Hence, as $k \to \infty$, the left hand side of inequality (7.1) tends to infinity while the right hand side goes to zero. This gives a contradiction in this case.

(ii) If $\|W_k A_k\|^2 - \|W_k(A_k + (Z_k\nabla A_k)^T \tilde{\tau}_k s_k)\|^2 \leq -\varepsilon_1$. Because $r_k \to \infty$ and $\tilde{\tau}_k \to 1$ as $k \to \infty$, we have

$$r_k\{\|W_k A_k\|^2 - \|W_k(A_k + (Z_k\nabla A_k)^T \tilde{\tau}_k s_k)\|^2\} \leq -r_k \varepsilon_1 \to -\infty. \tag{7.3}$$

Similar to the above case, $Pred_k \to -\infty$. This gives a contradiction with $Pred_k > 0$. These two contradictions prove the lemma. □

The following lemma studies the case when $r_k \to \infty$ as $k \to \infty$, and $\lim inf_{k\to\infty} \|W_k A_k\| = 0$.

Lemma 7.2. *Suppose GA_1-GA_3. If $r_k \to \infty$, as $k \to \infty$, and there exists a subsequence $\{k_j\}$ of the sequence of iterates that satisfies $\|W_k A_k\| > 0$ for all $k \in \{k_j\}$ and $\lim_{k_j\to\infty} \|W_{k_j} A_{k_j}\| = 0$, then a subsequence of the iteration sequence indexed $\{k_j\}$ satisfies Fritz John's conditions in the limit.*

Proof. Let the subsequence $\{k_i\}$ be renamed to $\{k\}$ to simplify the notations avoiding double indices. By using a contradiction we prove this lemma. So we presume that there exists no subsequence that satisfies the feasible Fritz John's conditions in the limit. By using lemma (5.3), there exists a constant ε_3 such that for all k sufficiently large,

$$\frac{|\ \|W_k A_k\|^2 - \|W_k(A_k + (Z_k\nabla A_k)^T \tilde{\tau}_k s_k)\|^2\ |}{\|W_k A_k\|^2} \geq \varepsilon_3. \tag{7.4}$$

We consider three cases:

(i) If $\lim inf_{k\to\infty} \frac{s_k}{\|W_k A_k\|} = 0$, the above inequality gives a contradiction.

(ii) If $\lim sup_{k\to\infty} \frac{s_k}{\|W_k A_k\|} = \infty$. From the way of computing s_k, we have

$$Z_k(\nabla f_k + r_k \nabla A_k W_k A_k) = -(B_k + \rho_k I)s,$$

where $\rho_k \geq 0$ is the Lagrange multiplier of the trust region constraint. Using the above equation, then inequality (6.1) can be written in the form

$$Pred_k \geq K_1 \tilde{\tau}_k \parallel Z_k(\nabla f_k + r_k \nabla A_k W_k A_k) \parallel \min \times \left\{ \delta_k, \frac{\parallel [\frac{1}{r_k} G_k + Z_k \nabla A_k W_k \nabla A_k^T Z_k + \frac{\rho_k}{r_k} I] s_k \parallel}{\parallel \frac{1}{r_k} G_k + Z_k \nabla A_k W_k \nabla A_k^T Z_k \parallel} \right\}. \tag{7.5}$$

Because $r_k \to \infty$, as $k \to \infty$, there exists an infinite number of acceptable steps such that inequality (7.1) holds. But inequality (7.1) can be written as

$$Pred_k < \|Z_k \nabla A_k\|^2 \|W_k A_k\|^2. \tag{7.6}$$

From inequalities (7.5) and (7.6), we have

$$K_1 \tilde{\tau}_k \parallel Z_k(\nabla f_k + r_k \nabla A_k W_k A_k) \parallel \min \times \left\{ \delta_k, \frac{\parallel [\frac{1}{r_k} G_k + Z_k \nabla A_k W_k \nabla A_k^T Z_k + \frac{\rho_k}{r_k} I] s_k \parallel}{\parallel \frac{1}{r_k} G_k + Z_k \nabla A_k W_k \nabla A_k^T Z_k \parallel} \right\} < b_1^2 \|W_k A_k\|^2,$$

where $b_1 = sup_{x\in\Omega} \|Z_k \nabla A_k\|$. Hence, if we divided the above inequality by $\|W_k A_k\|$, we obtain

$$K_1 \tilde{\tau}_k \parallel Z_k(\nabla f_k + r_k \nabla A_k W_k A_k) \parallel \min \times \left\{ \frac{\delta_k}{\|W_k A_k\|}, \frac{\parallel [\frac{1}{r_k} G_k + Z_k \nabla A_k W_k \nabla A_k^T Z_k + \frac{\rho_k}{r_k} I] s_k \parallel}{\parallel \frac{1}{r_k} G_k + Z_k \nabla A_k W_k \nabla A_k^T Z_k \parallel \|W_k A_k\|} \right\} < b_1^2 \|W_k A_k\|. \tag{7.7}$$

The right hand side of the above inequality goes to zero as $k \to \infty$. This implies that along the subsequence $\{k_i\}$ where $\lim_{k_i\to\infty} \frac{s_{k_i}}{\|W_{k_i} A_{k_i}\|} = \infty$, we have

$$\parallel Z_k(\nabla f_{k_i} + r_{k_i} \nabla A_{k_i} W_{k_i} A_{k_i}) \parallel \times \frac{\parallel [\frac{1}{r_{k_i}} G_{k_i} + Z_k \nabla A_{k_i} W_{k_i} \nabla A_{k_i}^T Z_k + \frac{\rho_{k_i}}{r_{k_i}} I] s_{k_i} \parallel}{\parallel \frac{1}{r_{k_i}} G_{k_i} + Z_k \nabla A_{k_i} W_{k_i} \nabla A_{k_i}^T Z_k \parallel \|W_{k_i} A_{k_i}\|},$$

is bounded. Therefore, asymptotically, either $\frac{s_{k_i}}{\|W_{k_i} A_{k_i}\|}$ lies in the null space of $Z_k \nabla A_{k_i} W_{k_i} \nabla A_{k_i}^T Z_k + \frac{\rho_{k_i}}{r_{k_i}} I$ or $\parallel Z_k(\nabla f_{k_i} + r_{k_i} \nabla A_{k_i} W_{k_i} A_{k_i}) \parallel \to 0$. The first possibility occurs only when $\frac{\rho_{k_i}}{r_{k_i}} \to 0$ as $k_i \to \infty$ and $\frac{s_{k_i}}{\|W_{k_i} A_{k_i}\|}$ lies in the null space of the matrix $Z_k \nabla A_{k_i} W_{k_i} \nabla A_{k_i}^T Z_k$ which contradicts assumption (7.4) and implies that a subsequence of the iteration sequence satisfies the Fritz John conditions in the limit. The second possibility implies that as $k_i \to \infty$, $\parallel Z_k(\nabla f_{k_i} + r_{k_i} \nabla A_{k_i} W_{k_i} A_{k_i}) \parallel \to 0$. Hence as $k_i \to \infty$, $r_{k_i} \|Z_k \nabla A_{k_i} W_{k_i} A_{k_i}\|$ must be bounded. This implies that a subsequence of the sequence of iterates satisfies the Fritz John conditions in the limit.

(iii) If $\lim sup_{k\to\infty} \frac{s_k}{\|W_k A_k\|} < \infty$ and $\lim inf_{k\to\infty} \frac{s_k}{\|W_k A_k\|} > 0$. Therefore $\|s_k\| \to 0$. Hence, as in the second case, the right hand side of (7.7) goes to zero as $k \to \infty$. This implies that

$$\| Z_k(\nabla f_k + r_k \nabla A_k W_k A_k) \| \frac{\| (Z_k \nabla A_k W_k \nabla A_k^T Z_k + \frac{\rho_k}{r_k} I) s_k \|}{\| Z_k \nabla A_k W_k \nabla A_k^T Z_k \| \, \|W_k A_k\|} \to 0.$$

But this implies that asymptotically, either $\parallel Z_k(\nabla f_k + r_k \nabla A_k W_k A_k) \parallel \to 0$ or $\frac{\|(Z_k \nabla A_k W_k \nabla A_k^T Z_k + \frac{\rho_k}{r_k} I) s_k\|}{\|Z_k \nabla A_k W_k \nabla A_k^T Z_k\| \|W_k A_k\|} \to 0$. As the second case, the two possibilities imply that a subsequence of the iteration sequence satisfies the Fritz John conditions in the limit. This completes the proof. □

8. Convergence when r_k is bounded

In this section, we presume that the parameter r_k is bounded. This means that, we presume the existence of an integer $\bar{k}$ such that for all $k \geq \bar{k}$, $r_k = \bar{r} < \infty$.

Lemma 8.1. *Suppose GA_1-GA_3. At any iteration indexed k at which $\|Z_k\nabla\phi(x_k;\bar{r}_k)\| + \|Z_k\nabla A_k W_k A_k)\| > \epsilon_1$, there exists a positive constant K_4 that depends on ϵ_1 but does not depend on k, such that*

$$Pred_k \geq K_4\tilde{\tau}_k\delta_k. \tag{8.1}$$

Proof. From equalities(2.13), (2.16), and general assumptions $GA_1 - GA_3$, then for all k, there exists $b_2 > 0$ such that $\|B_k\| \leq b_2$. Let $\|Z_k\nabla\phi(x_k;\bar{r}_k)\| > \frac{\epsilon_1}{2}$ and using inequality (6.1), we have

$$\begin{aligned} Pred_k &\geq K_1\tilde{\tau}_k \parallel Z_k\nabla\phi(x_k;\bar{r}_k) \parallel \min\left\{\delta_k, \frac{\parallel Z_k\nabla\phi(x_k;\bar{r}_k)\parallel}{\parallel B_k \parallel}\right\} \\ &\geq \frac{1}{2}K_1\tilde{\tau}_k\epsilon_1 \min\left\{1, \frac{\epsilon_1}{2b_2\delta_{max}}\right\}\delta_k \\ &\geq K_4\tilde{\tau}_k\delta_k, \end{aligned}$$

where $K_4 = \frac{1}{2}K_1\epsilon_1\min\{1, \frac{\epsilon_1}{2b_2\delta_{max}}\}$. □

Lemma 8.2. *Suppose GA_1-GA_3. If $\|Z_k\nabla\phi(x_k;\bar{r}_k)\| + \|Z_k\nabla A_k W_k A_k)\| > \epsilon_1$, then an acceptable step is found after finitely many trials i.e., the condition $\frac{Ared_{k_j}}{Pred_{k_j}} \geq \eta_1$ will be satisfied for some finite j.*

Proof. Since $\|Z_k\nabla\phi(x_k;\bar{r}_k)\| + \|Z_k\nabla A_k W_k A_k)\| > \epsilon_1$, then from lemmas (6.5) and (8.1), we have

$$\left|\frac{Ared_k}{Pred_k} - 1\right| = \frac{|\ Ared_k - Pred_k\ |}{Pred_k} \leq \frac{K_3\bar{r}\tilde{\tau}_k\delta_k^2}{K_4\tilde{\tau}_k\delta_k} \leq \frac{K_3\bar{r}\delta_k}{K_4}.$$

Now as the step s_{k_j} gets rejected, δ_{k_j} becomes small and eventually after finite number of trials, (i.e., for j finite), the acceptance rule will be met. This completes the proof. □

Lemma 8.3. *Suppose GA_1-GA_3. If $\|Z_k\nabla\phi(x_k;\bar{r}_k)\| + \|Z_k\nabla A_k W_k A_k)\| > \epsilon_1$, at a given iteration k, the jth trial step satisfies*

$$\|s_{k^j}\| \leq \frac{(1-\eta_1)K_4}{2\bar{r}K_3}, \tag{8.2}$$

then it must be accepted.

Proof. By using a contradiction we prove this lemma. Presume that the step s_{k^j} is rejected and inequality (8.2) holds. Then, from inequalities (6.7) and (8.1) we have

$$(1-\eta_1) < \frac{|\ Ared_{k^j} - Pred_{k^j}\ |}{Pred_{k^j}} < \frac{K_3\bar{r}\tilde{\tau}_{k^j}\|s_{k^j}\|^2}{K_4\tilde{\tau}_{k^j}\|s_{k^j}\|} \leq \frac{(1-\eta_1)}{2}.$$

This gives a contradiction and proves the lemma. □

9. Global convergence outcomes

In this section, we prove our master global convergence result for our trust-region algorithm.

Theorem 9.1. *Suppose GA_1–GA_3. Then the sequence of iterates generated by the algorithm satisfies*

$$\liminf_{k\to\infty}[\|Z_k\nabla f_k\| + \|Z_k\nabla A_k W_k A_k\|] = 0. \tag{9.1}$$

Proof. First, we prove that

$$\liminf_{k\to\infty}\|Z_k\nabla\phi(x_k;\bar{r}_k)\| + \|Z_k\nabla A_k W_k A_k\| = 0. \tag{9.2}$$

We prove (9.2) by contradiction. Suppose that, for all k, $\|Z_k\nabla\phi(x_k;\bar{r}_k)\| + \|Z_k\nabla A_k W_k A_k\| > \epsilon_1$. Let $k \geq \bar{k}$ and consider a trial step indexed j of the iteration indexed k such that $k^j \geq \bar{k}$. Using lemma (8.1), we have for any acceptable step indexed k^j,

$$\Phi_{k^j} - \Phi_{k^j+1} = Ared_{k^j} \geq \eta_1 Pred_{k^j} \geq \eta_1 K_4\tilde{\tau}_{k^j}\delta_{k^j}. \tag{9.3}$$

As k goes to infinity, then $\tilde{\tau}_{k^j} \to 1$ and the above inequality implies that

$$\lim_{k\to\infty}\delta_{k^j} = 0. \tag{9.4}$$

This implies that the radius of the trust region is not bounded below.

If we consider an iteration indexed $k^j > \bar{k}$ and if the previous step was accepted; *i.e.* if $j = 1$, then $\delta_{k^1} \geq \delta_{min}$. Hence δ_{k^j} is bounded in this case.

Now presume that $j > 1$. *i.e.*, there exists at least one rejected trial step. For the rejected trial step, we have from lemma (8.3)

$$\|s_{k^i}\| > \frac{(1-\eta_1)K_4}{2\bar{r}K_3},$$

for all $i = 1, 2, ... j-1$. Since s_{k^i} is a rejected trial step, then from the way of updating the radius of trust region (see algorithm (3.1)) and using the above inequality, we have

$$\delta_{k^j} = \alpha_1\|s_{k^{j-1}}\| > \alpha_1\frac{(1-\eta_1)K_4}{2\bar{r}K_3}.$$

Hence δ_{k^j} is bounded. But this contradicts (9.4). Therefore, the supposition is wrong. Hence,

$$\liminf_{k\to\infty}\|Z_k\nabla\phi(x_k;\bar{r}_k)\| + \|Z_k\nabla A_k W_k A_k\| = 0.$$

But this also implies (9.1). This completes the proof of the theorem. □

From the above theorem, we conclude that, given any $\epsilon_1 > 0$, the algorithm terminates because $\|Z_k\nabla f_k\| + \|Z_k\nabla A_k W_k A_k\| < \epsilon_1$.

10. Numerical outcomes

In this section, we present the numerical results of the interior-point trust-region Algorithm (3.3) which have been performed on a laptop with Intel Core (TM)i7-2670QM CPU 2.2 GHz and 8 GB RAM. Algorithm (3.3) was implemented as a MATLAB code and run under MATLAB version 7.10.0.499 (R2010a). A starting point $x_0 \in int(\boldsymbol{F})$ is given, we select $\delta_0 = max(\|s_0^{cp}\|, \delta_{min})$, where $\delta_{min} = 10^{-3}$, and we select $\delta_{max} = 10^5\delta_0$.

Table 10.1 Numerical outcomes of LANCELOT and proposed algorithm.

Problem Name	The number of variables	The number of equalities	The number of inequalities	LANCELOT iter(nfunc)	Propose algorithm iter(nfunc)
Problem 6	2	1	0	49(56)	15(25)
Problem 7	2	1	0	18(19)	9(12)
Problem 9	2	1	0	4(5)	16(17)
Problem 10	2	0	1	17(18)	26(29)
Problem 12	2	0	1	22(23)	8(9)
Problem 14	2	1	1	12(13)	11(12)
Problem 16	2	0	5	15(16)	3(4)
Problem 21	2	0	5	1(2)	3(4)
Problem 22	2	0	2	9(10)	22(24)
Problem 24	2	0	5	7(8)	33(41)
Problem 30	3	0	7	7(8)	3(4)
Problem 34	3	0	8	19(19)	26(27)
Problem 41	4	1	0	6(7)	20(21)
Problem 60	3	1	0	15(15)	11(13)
Problem 77	5	2	0	22(24)	21(22)
Problem 78	5	3	0	11(11)	10(15)
Problem 79	5	3	0	9(10)	5(12)

The values of the constants that are needed in Step 0 of algorithm (3.3) were select to be $\eta_1 = 0.25$, $\eta_2 = 0.75$, $\alpha_1 = 0.5$, $\alpha_2 = 2$, $\epsilon_1 = 10^{-7}$, and $\epsilon_2 = 10^{-6}$.

We report the numerical outcomes of the proposed algorithm in Table 10.1. The problems which are tested in this Table are the Hock and Schittkowski 's subset of the constrained and unconstrained testing environment [15].

In Table 10.1, we compare the numerical outcomes of algorithm (3.3) versus the corresponding outcomes of LANCELOT (Release A)[16].

The value of x_* and $f(x_*)$ are the same value indicated in Hock and Schittkowski [15]. In many of the test problems reported in Table 10.1, the number of iterations and the number of function evaluations of the interior-point trust-region algorithm are better than those obtained by LANCELOT. This indicates the viability of our approach.

11. Concluding remarks

We described a new interior-point penalty active-set trust-region algorithm for solving general nonlinear programming problem with bound on variables. The penalty method and the active set strategy are used in the proposed algorithm to transform the optimization problem with equality and inequality constraints with bound on variables to unconstrained optimization problem with bound on variables. The algorithm uses a Coleman–Li strategies and the requirement of strict feasibility to examine the optimality conditions for the bound constrained optimization problem.

There are many question should be answered for future work. We can improve the proposed algorithm to make it capable of treating nondifferentiation bound constrained optimization problem with equality and inequality constraints.

Acknowledgments

I am very grateful to the anonymous reviewers for their valuable and insightful comments, which have aided us in improving the quality of this paper. I would like to thank my professor Dr.El-Alem for his helpful comments.

References

[1] J. Dennis, M. El-Alem, K. Williamson, A trust-region approach to nonlinear systems of equalities and inequalities, SIAM J. Optim. 9 (1999) 291–315.

[2] B. El-Sobky, Arobust trust-region algorithm for general nonlinear constrained optimization problems, Department of Mathematics, Alexandria University, Alexandria, Egypt, 1998 Phd thesis.

[3] B. El-Sobky, A multiplier active trust-region algorithm for solving general nonlinear programming problem, Appl. Math. Comput. 219 (1) (2012) 928–946.

[4] B. El-Sobky, A penalty active-set trust-region algorithm for solving general nonlinear programming problem, Sylwan 9 (158) (2014) 273–290.

[5] A. Fiacco, G. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, John Wiley and Sons, New York, 1968.

[6] R. Fletcher, An l_1 penalty method for nonlinear constraints, In Numerical Optimization, SIAM Philadelphia, 1984. 26-40

[7] T. Coleman, Y. Li, An interior trust region approach for nonlinear minimization subject to bounds, SIAM J. Optim. 6 (1996) 418–445.

[8] I. Das, An interior point algorithm for the general nonlinear programming problem with trust region globalization. Technical Report 96-61, Institute for Computer Applications in Science and Engineering, NASA Langley Research Center Hampton, VA, USA, 1996.

[9] M. El-Alem, M.E. Sayed, B. El-Sobky, Local convergence of interior - point newton method for constrained optimization, J. Optim. Theory Appl. 120 (3) (2004) 487–502.

[10] Y. Yuan, Recent advances in trust region algorithms, Math. Program. Ser. B 151 (2015) 249–281.

[11] T. Steihaug, The conjugate gradient method and trust-region in large scale optimization, Siam J. Numer.Anal. 20 (1983) 3.

[12] Y. Yuan, On the convergence of a new trust region algorithm, Numer. Math. 70 (1995) 515–539.

[13] O. Mangasarian, Nonlinear Programming, McGraw-Hill Book Company, New York, 1969.

[14] M. El-Alem, A global convergence theory for the dennis, el-alem, maciel's class of trust- region algorithms for constrained optimization without assuming regularity, SIAM J. Optim. 9 (1999) 965–989.

[15] W. Hock, K. Schittkowski, Test Examples for Nonlinear Programming Codes, Lecture Notes in Economics and Mathematical Systems, vol. 187, Springer-Verlag, Berlin, 1981.

[16] A. Conn, N. Could, L. Toint, (LANCELOT): a Fortran package for Large-Scale Nonlinear (release A) Optimization. No. 17 in Springer Series in Computational Mathematics, Springer-Verlag, New York, 1992.